DOCUMENT RESUME

ED 041 701 RE 002 794

AUTHOR Hackett, Marie Gannon
TITLE A Hierarchy of Skills in Listening Comprehension and Reading Comprehension.
INSTITUTION Florida State Univ., Tallshassee. Dept. of Educational Research.
PUB DATE [70]
NOTE 27p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.45
DESCRIPTORS *Diagnostic Tests, Grade 2, Grade 5, Grade 8, Grade 11, *Listening Comprehension, Listening Skills, *Reading Comprehension, Reading Skills, *Sequential Learning, *Test Construction, Tests

ABSTRACT

An analysis was made of the data from a single criterion-referenced test which was constructed to measure a hierarchy of skills in listening and reading comprehension and which was administered to 1,186 subjects in grades 2, 5, 8, and 11. The research was concerned with the applicability of a hierarchically ordered achievement test to the diagnosis and assessment of listening and reading skills. Performance objectives were derived from 11 language comprehension skills and arranged in a hierarchical structure through the use of an heuristic procedural analysis. The design of the test was based on the theoretical constructs of Gagne. Paired items were scored on a yes/no basis and a percentage of probable response calculated. Correlations for paired test-retest listening scores ranged from .86 to .99 for the 11 skills. Paired test-retest reading score correlations ranged from .80 to .96. No serious bias was noted in the order of item presentation. Data were furnished on the development, validity and reliability, interpretations, and uses of the test. Tables and references are included. (WB)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED041701

A Hierarchy of Skills in Listening Comprehension And Reading Comprehension

Marie Gannon Hackett

A study was conducted to analyze the data from a criterion-referenced test which was constructed to measure a hierarchy of skills in listening comprehension and reading comprehension. The design of the test was derived from a theoretical position that a learning hierarchy, involving the notion of positive transfer of learning, represents a set of specified intellectual skills which are both ordered from more simple to more complex and which are also predicted to exhibit relationships compatible with the hypothesis of transfer from lower- to higher-level skills.

Eleven language comprehension skills were defined as performance objectives and arranged in a hierarchical structure through the use of an heuristic procedural analysis. Two parallel items were constructed to yield pass-fail information for each of the 11 skills. The content of these items was intended to sample situations relevant to each of the 11 skills, and in this sense to comprise a definition of these skills. Since the test was intended to measure the 11 skills at four grade levels, an attempt was made to control the

RE002 794

intrinsic difficulty of items by means of appropriate language complexity, vocabulary, and interest.

The test was administered to 1,186 subjects at the second, fifth, eighth, and eleventh grades. Results for the four grades studied may be summarized relative to research questions that were formulated to investigate the 11 cognitive skills and their associated performance effects. Differences were obtained in the degree of the consistency of measures of listening and reading, as shown by pass-fail measures on two items for each skill, with a high proportion reaching a level of consistency of .70. Correlations of .86, .86, .92, .98, .88, .96, .94, .66, .65, and .99 for each of the 11 skills, respectively, were found for paired test-retest listening scores. Correlations of .80, .88, .90, .96, .98, .90, .95, .92, .72, .78, .96 for each of the 11 skills respectively, were found and paired test-retest reading scores. Correlations of .97, .95, .92, and .95 for grades two, five, eight, and eleven, respectively, were found between an original testing order and a scrambled testing order, indicating no serious bias arising from this variable.

Conditional probabilities of correct responses to listening measures indicated ordered patterns of predictable relationships

of lower-level to higher-level skills. Scalogram analyses yielded reproducibility coefficients of .88, .86, .80, and .83 for grades two, five, eight, and eleven, respectively, indicating the extent to which test scores can be predicted to fit the model of ordered patterns of difficulty.

In conclusion, a study was conducted by administering a test with four levels in oral and printed form to subjects in grades two, five, eight, and eleven. This research was primarily concerned with the investigation, according to theoretical considerations, of a hierarchy of intellectual skills in listening comprehension and reading comprehension. Data were also furnished on the development, validity and reliability, interpretations, and uses of the test.

A Hierarchy of Skills in Listening Comprehension and Reading Comprehension

By

Marie Gannon Hackett

Department of Educational Research
Florida State University
Tallahassee 32306

A study was conducted to analyze the data from a single criterion-referenced test which was constructed to measure a hierarchy of skills in listening comprehension and reading comprehension, administered to 1, 186 subjects at the second, fifth, eighth, and eleventh grades. Based upon specific language skills which were identified and defined as performance objectives, the test was also designed to discover the sequence and the predictable relationships among the comprehension skills. Undoubtedly many factors contribute to the learning of such skills. However, this research was primarily concerned with the applicability of theoretical constructs proposed by Gagné (1965) to the measurement of the initial capabilities of the learner in relation to the order and dependence of listening comprehension skills and of reading comprehension skills.

The design of the test was derived from a theoretical position that a learning hierarchy, involving the notion of positve transfer of learning, represents a set of specified intellectual skills which are both ordered from more simple to more complex and which are also predicted to exhibit relationships compatible with the hypothesis of transfer from lower-to higher-level skills. Specifically, this research was concerned with the applicability of a hierarchically-ordered test of achievement to the diagnosis and assessment of listening skills and reading skills.

The problem of identifying and measuring the effects of prior learning on performance of listening skills and reading skills is related to the

the question of sequence of skills and positive transfer among the skills. In a recent review of studies concerned with conditions for instructional psychology, Gagné & Rohwer (1969) noted that there is considerable evidence for the notion that the learning of particular classes of tasks depends in a positive transfer sense on the prior learning of other particular classes of performance. Specifically, the authors concluded that learning verbal associations typically receives much positive transfer from prior discrimination learning, stimulus coding, and response integration; concept learning from prior learning on dimension discrimination; rule learning from prior concept learning; and problem solving from prior learning of relevant rules. Gagné's (1965) hypothesis that certain kinds of learning are necessary prerequisites, i.e., transfer positively, to other kinds of learning suggests that the hierarchical nature of learning tasks is one of the critical conditions of learning complex performances.

Several studies have attempted to identify and analyze the hierarchical processes involved in learning tasks (Gagné & Paradise, 1961, Gagné, 1962, Gagné, et. al., 1965, Gibson, 1965, Schutz, Baker, & Gerlach, 1965, and Cox & Graham, 1966). Recent studies in the category of concept learning and rule learning are concerned with identifying kinds of prior learning which contribute to, i.e., transfer positively to, the learning of a given class of performance. (Marchbanks & Leven, 1965, McNeil & Stone, 1965, Samuels & Jeffrey, 1966, Kingsley & Hall, 1967, and Beilin, Kagan, & Rabinowitz, 1966).

Although there has been considerable research upon the question of sequence of skills and positive transfer among the skills, relatively little research has been conducted relating to the identification of the sequence and transfer among the language comprehension skills of listening and reading. Referring in a recent review of research to serious questions concerning what is known about listening, Devine (1967) stated that studies of measurement are needed to support assumptions about the listening process. Davis (1967) noted that there has been a surprisingly small number of experimental studies in comprehension despite the long standing interest in reading as a thought process. Observing that standardized reading tests often mask some of the important outcomes of instruction because they measure a conglomerate of skills and abilities at the same time, Chall (1967) pointed to the need for single component tests of skills, particularly of reading comprehension skills.

The theoretical setting for the present study involved the idea that certain language comprehension skills might be analyzed as intellectual strategies which identify and define the processes in listening comprehension and in reading comprehension. Specifically, intellectual skills are distinguishable as hierarchical classes of component skills on the basis of different outcome performances. It was theorized that an analysis of the language comprehension process would identify a hierarchy of skills requiring rule-using behaviors and problem-solving behaviors

which could be measured as intellectual skills and which are mediators of positive transfer among themselves, ordered from more simple to more complex. However, it should be emphasized that both rule-using behavior and also problem-solving behavior require that an individual possesses prerequisite capabilities, e.g., previously learned concepts.

Procedure

A review of the literature yielded a list of the most important listening and reading comprehension skills. It was desired that such skills be operationally defined and measured as performance objectives (Gagné, 1964; Mager, 1962; Tyler, 1951). Skills were sought which would require rule-using behaviors and problem-solving behaviors for their successful performance. Since investigators of both listening comprehension and reading comprehension appeared to emphasize quite similar intellectual skills as necessary in the language comprehension process, a single test was designed to measure each of the skills in two forms: oral and printed.

Using the list of skills expressed as performance objectives, it appeared logical to construct a hypothetical hierarchy of skills by attempting to answer the question suggested by Gagné's (1962) procedural analysis for arranging skills in a hierarchical structure: What would an individual have to be able to do in order to perform the final objective, e.g., listening comprehension or reading comprehension? The

research question implied in such an analysis is that if an individual is able to perform successfully a skill within the hierarchy, he should also be able to perform successfully more simple skills lower in the hierarchy. Subject to empirical findings a hierarchy of language comprehension skills from more simple to more complex was identified and defined as follows: (1) identifying the stated main idea; (2) providing examples by detail; (3) reinstating a sequence of ideas; (4) inferring the main idea from specifics; (5) identifying mood; (6) applying standards to judge persuasion; (7) predicting the sequence of thought; (8) inferring connotative word meaning; (9) identifying sequence ambiguities; (10) inferring speaker's or writer's purpose; (11) judging logical validity.

Specifically, if an individual is able to infer a main idea from specifics in a passage, it appeared reasonable to assume that he would also be able to perform more simple skills, skills lower in the hierarchy, e.g., reinstating a sequence of ideas, providing examples of details, and identifying the stated main idea. In addition, the investigator hypothesized that successfully judging the logical validity of a passage probably indicated that an individual possessed lower-level skills in the hierarchy. Empirical data were to be sought in order to investigate the hypothesized order of the skills.

In order to measure an individual's ability to demonstrate performance of each of the comprehension skills, two parallel items for each

of the comprehension skills, two parallel items for each of the eleven skills were constructed to yield pass-fail information on each of the skills. The content of the items was intended to sample the situations about which conclusions could be drawn relative to criterion performance. Thus, the content of the test may be regarded as an explication of the eleven previously defined performance objectives. To eliminate spurious interrelationships among skill scores for items based on the same passage, each of the items was based on a different passage.

The results of this study necessarily depend on the content validity of the items used. No statistical manipulation of data resulting from use of items lacking intrinsic validity can wholly make up for their fundamental inadequacy (Davis, 1967). Empirical findings relative to item performance were expected either to confirm the appropriateness of the items or to suggest revisions of the items.

Since the test was intended to measure the 11 skills at four grade levels--two, five, eight, and eleven--questions relative to the skills which followed the passages were similar across the grade levels. Avoiding difficult misleads, multiple choice responses were constructed to measure the skills, since it was expected that empirical evidence would demonstrate that items had performed as intended. Furthermore, in an attempt to control intrinsic difficulty of content, items were constructed to meet criteria of appropriate difficulty: (1) language complexity--

coordination and subordination determined by length and number of sentences, (2) vocabulary--selected from word lists and teachers' opinions, and (3) interest--judged by basal readers, literature texts, and teachers' opinions.

Passages containing few, short sentences for grade two, many, short sentences for grade five, few, long sentences for grade eight, and many, long sentences for grade eleven were constructed. A few sentences ranged from six to nine. Short sentences included five to twelve words and long sentences included eight to twenty-five words. Vocabulary and interest were selected and judged appropriate from texts commonly found in the schools of California and word lists, e.g., Lorge & Thorndike (1944). These sources were reviewed by 28 teachers at the various grade levels on two occasions for interest and acceptability.

The test consisted of four levels, A, B, C, and D for grades two, five, eight, and eleven, respectively. Subjects in each class, having been randomly assigned to one of two testing conditions, took the test in oral form--the Listening Test--and in printed form--the Reading Test--in separate class sessions. Each of two parallel items was scored pass-fail and performance of each skill was scored (1) pass, if both items had been scored pass, and (2) fail, if one or both items had been scored fail, making the maximum possible score 11.

Results

Since each item pair was intended to measure a different skill, a measure of reliability of the scoring procedure for the test seemed to be unavailable from standard statistics. A test of the degree to which "inconsistency" (+- or -+) differed from "consistency" (++ or --) in the measure of each skill more than would be expected by chance was needed. The percent consistency scores, with a high proportion reaching a consistency index of .70, were obtained by adding the number of ++ and -- responses and dividing by the total. Correlations of .86, .86, .92, .98, .98, .88, .96, .94, .66, .65, and .99 for each of the 11 skills, respectively, were found for paired test-retest listening scores. Correlations of .80, .88, .90, .96, .98, .90, .95, .92, .72, .78, .96 for each of the 11 skills respectively, were found for paired test-retest reading scores.

Evidence from the study of listening skills will be reported relating to the existence of a hierarchy, although data have also been collected and analyzed from the Reading Test. The listening skills were ranked by difficulty level in terms of the ordered probabilities, i.e., averages of correct (pass) responses. Difficulty levels are summarized in Table 1. For the second grade subjects, the skill with the greatest probability of correct responses is Skill 4; the skill with the second highest probability of responses is Skill 5; and the skill with the lowest probability of correct responses is Skill 9. Inspection of the rankings across the

TABLE 1

DIFFICULTY LEVEL BY RANKS 1-11 FOR SKILLS BASED ON PROBABILITIES AND AVERAGES OF CORRECT RESPONSES FOR LISTENING MEASURES AT GRADES TWO, FIVE, EIGHT, AND ELEVEN

Skill	Skill Names	Grade 2 n = 84		Grade 5 n - 265		Grade 8 n = 124		Grade 11 n = 96	
		Difficulty Level	Per Cent Passing	Difficulty Level	Per Cent Passing	Difficulty Level	Per Cent Passing	Difficulty Level	Per Cent Passing
1	Identifying stated main idea	5	.61	8	.28	10	.33	2	.67
2	Recognizing examples by detail	3	.76	5	.59	1	.83	1	.69
3	Reinstating sequence of ideas	6	.54	4	.68	7	.46	7	.33
4	Inferring main idea from specifics	1	.88	10	.16	11	.19	3	.54
5	Identifying mood	2	.82	1	.86	8	.44	4	.51
6	Applying standards to judge persuasion	9	.18	6	.36	5	.53	11	.14
7	Predicting sequences of thought	8	.21	3	.72	3	.61	8	.26
8	Inferring connotative word meaning	4	.62	2	.75	2	.74	9	26
9	Identifying sequence inconsistencies	11	.02	11	.14	9	.39	6	.45
10	Inferring speaker's purpose	7	.52	9	.22	4	.60	5	.47
11	Judging logical validity	10	.04	7	.33	6	.48	10	.24

four levels of difficulty for the skills differed at grade levels.

Although it would seem possible to obtain an average ranking of difficulty for each skill at the four grade levels, the per cent consistency scores in the measure of each skill suggested that some items were not adequately measuring the skill. Therefore, an attempt was made not only to inspect difficulty levels for skills for each grade but also to further analyze the nature of the skills.

Since an attempt had been made to arrange the skills in a hierarchical order from more simple to more complex, it was considered important to investigate possible bias in the effects of testing order. Correct responses might depend on position order rather than on order of difficulty. Therefore, it seemed reasonable to test the additional hypothesis of the extent to which ordering was a function of testing. However, when items in the test were presented in scrambled order to a group of students, the difficulty level by ranks of the skills was not significantly different from the original order of presentation. In Table 2 the results are shown for the scrambled order of the test. Since the data were available in terms of rank orders, Spearman's rank-difference method was applied and yielded a correlation between the two rankings of .97, .95, .92, and .95, respectively.

If a hierarchy of skills exists, the conditional probabilities of correct responses should indicate the degree to which predictable relationships

TABLE 2

DIFFICULTY LEVEL BY RANKS 1-11 FOR SKILLS PRESENTED IN SCRAMBLED ORDER BASED ON PROBABILITIES OF CORRECT RESPONSES FOR LISTENING MEASURES AT GRADES TWO, FIVE, EIGHT, AND ELEVEN

Skill	Skill Names	Grade 2 n = 84	Grade 5 n = 265	Grade 8 n = 124	Grade 11 n = 96
1	Identifying stated main idea	5	8	8	4
2	Recognizing examples by detail	3	5	1	1
3	Reinstating sequence of ideas	6	4	10	8
4	Inferring main idea from specifics	1	10	11	2
5	Identifying mood	2	1	9	3
6	Applying standards to judge persuasion	8	7	9	3
7	Predicting sequences of thought	10	3	3	7
8	Inferring conotative word meaning	4	2	2	10
9	Identifying sequence inconsistencies	11	9	7	6
10	Inferring speaker's purpose	7	11	4	5
11	Judging logical validity	9	6	6	9

obtain among the skills. Table 3 presents the probability of obtaining a correct response to Skill X_2 given that Skill X_1 is mastered. The probabilities shown in Table 6 represent the degree to which the attainment of one skill can be predicted from the attainment of another skill. For example, inspection of Table 6 shows that the probability for success on Skill 5, given that Skill 4 had been achieved, was .82; the probability for success on Skill 2, given that Skill 4 had been achieved, was .73; but the probability for success on Skill 9, given that Skill 4 had been achieved, was only .03. Thus, weaker relationships of predictability are seen between skills which are higher in the hierarchy.

Since meaningful probabilities are predicted when Skill X_1 precedes Skill X_2 in the hierarchy, the values lying above the diagonal will yield additional significant information regarding successive pairs of lower- and higher-level skills. Specifically, if Skill 4 was achieved, the probability of success on Skill 5 was .82; if Skill 5 was achieved, the probability of success on Skill 8 was .67. It may be noted that the values of conditional probability shown here are not measures of difficulty; and that these values may vary independly of the difficulty measure applicable to any given skill. The predictability of a lower-level to higher-level relationship might actually be .00 (and it may be noted that some few such values were obtained) without influencing the difficulty measure <u>per se</u>. Thus, the values of conditional probability ranging up to .82 indicate high degrees of predictability.

TABLE 3

PROBABILITY OF OBTAINING A CORRECT RESPONSE TO SKILL X_2 GIVEN THAT SKILL X_1 IS MASTERED FOR SECOND GRADE SUBJECTS ON THE LISTENING MEASURE (N = 84)

Skill X_2	4	5	2	8	1	3	6	7	6	11	9
X_1											
4	---	.82	.73	.62	.62	.55	.53	.22	.19	.03	.03
5	.88	---	.78	.65	.62	.55	.55	.25	.20	.03	.03
2	.84	.84	---	.67	.61	.53	.52	.23	.20	.03	.03
8	.89	.87	.83	---	.71	.54	.63	.31	.21	.04	.04
1	.90	.84	.77	.73	---	.67	.59	.28	.22	.04	.04
3	.91	.84	.76	.62	.76	---	.60	.22	.22	.07	.04
10	.89	.86	.75	.73	.68	.61	---	.30	.16	.05	.02
7	.89	.94	.83	.89	.78	.56	.72	---	.17	.06	.06
6	.93	.93	.87	.73	.73	.67	.47	.20	---	.13	.00
11	.67	.67	.67	.67	.67	1.00	.67	.33	.67	---	.00
9	1.00	1.00	1.00	1.00	1.00	1.00	.50	.50	.00	.00	---
Median	.89	.84	.78	.70	.69	.59	.57	.27	.20	.04	.03

THE MARGINAL LEGIBILITY OF THIS PAGE IS DUE TO POOR ORIGINAL COPY. BETTER COPY WAS NOT AVAILABLE AT THE TIME OF FILMING. E.D.R.S.

As skills get higher in the hierarchy, the probability of getting them correct decreases. Specifically, the decreasing probability of correct responses to higher-level skills shows, in general, that as skills become more widely separated in the hierarchy, the amount of transfer among them diminished. Consequently, the patterns of conditional probabilities of correct responses indicate that success on any specified skill is predictable from success on skills lower in the hierarchy. The conditional probabilities of correct responses of the skills for grades five, eight, and eleven are given in Tables 4, 5, and 6.

Finally, in order to obtain another measure of ordered relationships of the skills, the data were analyzed using the Guttman Scalogram Analysis technique. Ranking scores from highest to lowest, and ranking skills from most favorable to least favorable, subjects with the highest scores--highest being most favorable--would have answered only the most favorable items; those scoring low would have answered only the least favorable items. The analysis yields a coefficient of reproducibility which indicates how well an individual's response pattern can be predicted knowing his total score. The evidence from such an analysis appeared to support further the theoretical prediction that: (1) a particular skill might transfer positively to an adjacent skill in the hierarchy and (2) the successful performance of a skill would insure the successful performance of subordinate skills in the hierarchy. Reproducibility coefficients for each of the grades studied are presented in Table 7. Although

TABLE 4

PROBABILITY OF OBTAINING A CORRECT RESPONSE TO SKILL X_2 GIVEN THAT SKILL X_1 IS MASTERED FOR FIFTH GRADE SUBJECTS ON THE LISTENING MEASURE (N = 265)

Skill X_2 / X_1	5	8	7	3	2	6	11	1	10	4	9
5	---	.81	.79	.74	.65	.40	.34	.30	.24	.18	.15
8	.93	---	.78	.75	.68	.39	.36	.30	.25	.19	.15
7	.95	.82	---	.82	.68	.40	.35	.32	.25	.17	.15
3	.93	.82	.86	---	.71	.36	.37	.33	.25	.19	.16
2	.94	.87	.82	.82	---	.37	.35	.33	.26	.19	.12
6	.98	.83	.80	.69	.62	---	.37	.27	.27	.18	.16
11	.90	.83	.77	.77	.64	.41	---	.35	.35	.20	.56
1	.93	.82	.84	.82	.71	.34	.41	---	.22	.22	.18
10	.95	.86	.83	.79	.72	.44	.53	.28	---	.16	.25
4	1.00	.90	.81	.83	.73	.42	.42	.39	.22	---	.10
9	.94	.81	.78	.78	.53	.42	.23	.36	.39	.11	---
Median	.94	.83	.81	.79	.68	.40	.37	.33	.25	.19	.16

TABLE 5

PROBABILITY OF OBTAINING A CORRECT RESPONSE TO SKILL X_2 GIVEN THAT SKILL X_1 IS MASTERED FOR EIGHTH GRADE SUBJECTS ON THE LISTENING MEASURE (N = 129)

Skill X_2	2	7	7	10	6	11	3	5	9	1	4
X_1											
2	---	.79	.64	.62	.60	.52	.50	.48	.45	.34	.19
8	.88	---	.66	.67	.58	.52	.50	.51	.48	.33	.22
7	.87	.80	---	.63	.57	.51	.50	.49	.45	.38	.24
10	.87	.84	.65	---	.54	.54	.50	.46	.45	.35	.24
6	.94	.80	.65	.61	---	.58	.58	.50	.47	.39	.18
11	.88	.80	.65	.67	.63	---	.58	.50	.42	.33	.23
3	.90	.81	.67	.65	.67	.61	---	.44	.42	.33	.23
5	.91	.87	.69	.63	.61	.56	.46	---	.50	.39	.20
9	.96	.92	.71	.69	.65	.52	.50	.56	---	.48	.23
1	.85	.73	.71	.63	.63	.49	.46	.51	.56	---	.15
4	.83	.83	.75	.75	.50	.58	.54	.46	.46	.25	---
Median	.88	.81	.67	.64	.61	.53	.50	.49	.46	.36	.23

TABLE 6

PROBABILITY OF OBTAINING A CORRECT RESPONSE TO SKILL X_2 GIVEN THAT SKILL X_1 IS MASTERED FOR ELEVENTH GRADE SUBJECTS ON THE LISTENING MEASURE (N = 96)

Skill X_2	2	1	4	5	10	9	3	7	8	11	6
X_1											
2	---	.68	.59	.58	.46	.46	.36	.29	.26	.26	.15
1	.70	---	.53	.53	.66	.63	.41	.38	.38	.34	.20
4	.75	.65	---	.58	.50	.44	.37	.29	.25	.33	.11
5	.78	.69	.61	---	.53	.53	.43	.27	.31	.20	.15
10	.67	.93	.58	.58	---	.67	.40	.42	.38	.36	.22
9	.70	.93	.53	.61	.70	---	.47	.30	.37	.33	.16
3	.75	.81	.59	.66	.56	.63	---	.38	.28	.38	.22
7	.76	.96	.60	.52	.76	.52	.48	---	.48	.48	.36
8	.68	.96	.52	.60	.68	.64	.36	.48	---	.36	.24
11	.74	.96	.74	.44	.70	.61	.52	.52	.39	---	.30
6	.77	1.00	.39	.54	.77	.54	.54	.69	.46	.54	---
Median	.75	.93	.59	.58	.67	.58	.42	.38	.38	.35	.21

TABLE 7

REPRODUCIBILITY COEFFICIENTS OF LISTENING MEASURES WITH SUBJECTS RESPONDING AT GRADES TWO, FIVE, EIGHT, AND ELEVEN

Grade	Reproducibility Coefficient
2 (n = 84)	.88
5 (n = 265)	.86
8 (n = 124)	.80
11 (n = 96)	.83

Guttman (1944) suggested that a reproducibility coefficient value of .90 is an acceptable lower limit. Edwards (1957) indicated that a vlue of .85 shows the per cent accuracy with which responses to various statements can be reproduced from total scores. Since perfect scales exist only as ideal models, it is particularly useful to determine the extent to which success or failure for subjects with known test scores can be predicted to fit the model of an ordered scale.

Discussion

Data were obtained from a criterion-referenced test of listening comprehension and reading comprehension which defined and specified 11 skills to be measured. Definitions were specified in terms of human performance criteria. Each item was designed to measure a particular class of performance. Thus the basic form of measurement was pass or fail. Two items, measuring a class of performance or skill, seemed sufficient, since the operation of unknown factors, varying randomly in their effects, would not be expected to occur twice in exactly the same way.

Validity in the sense of representativeness of what was measured (content validity) was a critical consideration in the design of each item. Thus, the content of the test may be regarded as an explication of the eleven defined performance objectives.

Concerning the stability of test measures, differences were obtained as an index of consistency for measures of listening and reading, as

shown by pass-fail measures on two items for each skill, with a high proportion reaching a level of consistency of .70. As to the reliability of test measures, correlations of .86, .86, .92, .98, .98, .88, .96, .94, .66, .65, and .99 for each of the 11 skills, respectively, were found for paired test-retest listening scores. Correlations of .80, .88, .90, .96, .98, .90, .95, .92, .72, .78, and .96 for each of the 11 skills respectively, were found for paired test-retest reading scores. In the matter of the possible effects of testing order, correlations of .97, .95, .92, and .95 for grades two, five, eight, and eleven, respectively, were found between an original testing order and a scrambled testing order of listening measures, indicating no serious bias arising from this variable.

With respect to the sequence and transfer among the skills, conditional probabilities of correct responses to listening measures indicated patterns of predictable relationships of lower-level to higher-level skills at all four grade levels studied. Finally, in a measure obtained to show the extent of the relationships among the skills, scalogram analyses yielded reproducibility coefficients of .88, .86, .80, and .83 for grades two, five, eight, and eleven, respectively, indicating the extent to which test scores can be predicted to fit the model of ordered patterns of difficulty.

The results were analyzed in order to study the characteristics of the test. Since a measure of reliability of the scoring procedure seemed

unavailable, per cent consistency scores were computed. Although a reasonably high proportion of pass-fail measures on two items for each skill reached a consistency index of .70, two items can hardly be expected to provide a stable indication of performance. A revised test, used in a current study, included not only the construction of a third item to measure each of the skills but also the modification of the scoring procedure for certain items and the alternation of items whose per cent consistency was less than .70. However, it should be noted that the skills were shown to have decidedly satisfactory test-retest values.

The theoretical assumption that the skills would exhibit relationships of ordering compatible with the hypothesis of transfer from lower-levels to higher-levels was given support from evidence from (1) an analysis of the conditional probabilities of correct responses, and (2) a Guttman-type analysis. Although the results from these analyses were highly encouraging, more refined analyses are needed to provide improved prediction. The results highlight the need for analytic studies of language skills which would verify a specific causal relationship among the skills here identified as "higher" and "lower," as well as the possibility of studies of transfer of learning from other subordinate skills not yet identified.

REFERENCES

Beilin, H., Kagan, J., Rabinowitz, R. Effects of verbal and perceptual training on water level representation. Child Development, 1966, 37, 317-330.

Chall, J. S. Learning to read: The great debate. New York: McGraw Hill, 1967.

Cox, R. C. & Graham, G. T. The development of a sequentially scaled achievement test. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1966.

Davis, F. B. Identification and Measurement of Reading Skills of High School Students. Cooperative Research Project No. 3023. University of Pennsylvania, Philadelphia, 1967.

Devine, T. G. Listening. Review of Educational Research, 1967, 37, 152-158.

Edwards, A. L. Techniques of attitude scale construction. New York: Appleton-Century Crofts, 1956.

Gagné, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagné, R. M. The implications of instructional objectives for learning. In C. M. Lindvall (Ed.), Defining educational objectives. Pittsburgh, Pa.: University of Pittsburgh Press, 1964.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart & Winston, 1965.

Gagné, R. M. & Paradise, N. E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, 1961, 75. (Whole No. 518).

Gagné, R. M. & Rohwer, Jr., W. D. Instructional Psychology. Annual Review of Psychology, 1969, 20, in press.

Gagné, R. M. & Staff, Univ. of Md. Mathematics Project. Some factors in learning nonmetric geometry. Monograph Social Research In Child-Development, 1965, 30 (1), 42-49.

Gibson, E. J. Learning to read. Science, 1965, 148, 1066, 1072.

Guttman, L. A basis for scaling qualitative data. The American Sociological Review, 1944, 9, 139-150.

Hively, W., Patterson, H. L., & Page, S. H. Generalizability of performance by Job Corps trainees on a universe-defined system of achievement tests in elementary mathematical calculation. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1968.

Kingsley, R. C. & Hall, V. C. Training conservation through the use of learning sets. Child Development, 1969, 38, 1111-1126.

Mager, R. F. Preparing objectives for programmed instruction. San Francisco: Fearon, 1962.

Marchbanks, G. & Levin, H. Cues by which children recognize words. Journal of Educational Psychology, 1965, 56, 57-61.

McNeil, J. D. & Stone, J. Note on teaching children to hear separate sounds in spoken words. Journal of Educational Psychology, 1965, 56, 13-15.

Samuels, S. J. & Jeffrey, W. E. Discriminability of words and letter cues used in learning to read. Journal of Educational Psychology, 1966, 56, 337-340.

Schutz, R. E., Baker, R. L., & Gerlach, V. S. Measurement procedures in programmed instruction. Tempe, Ariz.: Classroom Learning Laboratory, Arizona State University, 1964.

Thorndike, E. L. & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College, Columbia University, 1944.

Tyler, R. W. The functions of measurement in improving instruction. In E. F. Lindquist, Educational Measurement. Washington, D. C.: American Council on Education, 1951.